# Computational Methods of Optimization
# First Midterm(22nd Sep, 2019)

**Instructions:**

- This is a closed book test. Please do not consult any additional material.
- Attempt all questions
- Total time is 90 mins.
- Answer the questions in the spaces provided. Answers outside the spaces provided will not be graded.
- Rough work can be done in the spaces provided at the end of the booklet

Name: ______________________________

SRNO: Degree: Dept:

| Question: | 1 | 2 | 3 | 4 | 5 | 6 | Total |
|---|---|---|---|---|---|---|---|
| Points: | 10 | 10 | 5 | 10 | 10 | 5 | 50 |
| Score: | | | | | | | |

In the following, assume that $f$ is a $\mathcal{C}^1$ function defined from $\mathbb{R}^d \to \mathbb{R}$ unless otherwise mentioned. Also $\mathbf{x} = [x_1, x_2, \ldots, x_d]^\top \in \mathbb{R}^d$ and $\|\mathbf{x}\| = \sqrt{\mathbf{x}^\top \mathbf{x}}$. Set of real symmetric $d \times d$ matrices will be denoted by $\mathcal{S}_d$. $[n]$ will denote the set $\{1, 2, \ldots, n\}$

1. (10 points) Please indicate True(T) or False(F) in the space given after each question. All questions carry equal marks
   (a) Let $a < b$ where $a, b \in \mathbb{R}$ and $h : [a, b] \to \mathbb{R}$ be differentiable and satisfies $h(a) = h(b)$. Then $h$ has a critical point in $(a, b)$. Recall that a critical point is a point, $\mathbf{x}$, such that $\nabla f(\mathbf{x}) = 0$. __**T**__
   (b) In the previous question let $h(x) = \alpha_1 x^2 + \alpha_2 x + \alpha_3$The values of $\alpha_1, \alpha_2, \alpha_3$ are not given but it is given that $h(a) = h(b) = 0$ and it is given that $h(x) > 0$ for all $x \in (a, b)$. The function $h$ is convex. __**F**__
   (c) If $f$ is a coercive function bounded from below then the global mimimum must lie at one of the critical points. __**T**__
   (d) Consider $g : \mathbb{R} \to \mathbb{R}, g(u) = \frac{1}{2}u^2 - \frac{1}{3}u^3$.The function has a global minimum. __**F**__
   (e) The point $u = 0$ is a local maximum of $g$, defined in the previous question. __**F**__
   (f) If all critical points of $f$ are global minima then the function must be convex. __**F**__
   (g) The Hessian of $f$ is positive definite everywhere. The cardinality of the set of critical points of $f$ is three? __**F**__
   (h) Let $f : \mathbb{R}^2 \to \mathbb{R}$. The Hessian at a critical point $\mathbf{x}$ is $H(\mathbf{x}) = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}$
   (i) The set $\{(\mathbf{x}, t) | \|\mathbf{x}\| \le t\}$ is convex. __**T**__
   (j) Let $\{\mathbf{x} | A\mathbf{x} = \mathbf{b}\}$ is not a convex set. __**F**__

2. Consider minimization of the function $f : \mathbb{R}^3 \to \mathbb{R}$ defined as follows

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top A\mathbf{x} - \mathbf{b}^\top \mathbf{x}$$

Where $A = \begin{bmatrix} 1 & a & b \\ -a & 2 & c \\ -b & -c & z \end{bmatrix}$ and $\mathbf{b} = [1, 2, 1]^\top$. The values of $a, b, c \in \mathbb{R}$ are not known.

   (a) (4 points) From this information is it possible to determine the gradient and Hessian of $f$ at $\mathbf{x} = [1, 1, 1]^\top$ assuming that $z = 1$? If not what minimal additional information is required to evalute the gradient and Hessian.

**Solution:** Yes it is possible.

$$\nabla f(\mathbf{x}) = \frac{1}{2}(A + A^\top)\mathbf{x} - \mathbf{b}, \quad \nabla f(0) = [1, 2, 1]^\top - \mathbf{b} = 0$$

$$H(\mathbf{x}) = \frac{1}{2}(A + A^\top) = Diag(1, 2, z)$$

where $Diag(v)$ is a diagonal matrix with $i$th diagonal entry to be substituted by $v_i$.

   (b) (3 points) Determine if the global minimum exist at $z = 1$? If not give reasons. If yes compute it.

**Solution:** For z=1, the function is convex as the Hessian is P.d. and the global minimum is attained at $\mathbf{x} = [1, 1, 1]^\top$.

(c) (3 points) Repeat the above two questions for $z = -1$. if we want to determine global maxima.

**Solution:** For $z = -1$ neither the global minimum or global maximum is attained.

3. (5 points) If $f$ is convex function defined over $C \subset \mathbb{R}^d$ then for any $\beta \in \Delta_n$ where $\Delta_n = \{\gamma \in \mathbb{R}^n | \sum_{i=1}^n \gamma_i = 1, \gamma_i \geq 0, i \in [n]\}$

$$f(\sum_{i=1}^n \beta_i \mathbf{x}_i) \leq \sum_{i=1}^n \beta_i f(\mathbf{x}_i)$$

holds for any $\mathbf{x}_1, \ldots, \mathbf{x}_n \in C$. If $f$ is strictly convex then equality is attained only whem $\mathbf{x}_i = \mathbf{x}$ for all $i \in [n]$. Show that for any $x_i > 0, i \in [n], x_i \in \mathbb{R}$ and $\gamma \in \Delta_n$

$$\prod x_i^{\gamma_i} \leq \sum_{i=1}^n \gamma_i x_i$$

**Solution:** Let $f(z) = -log(z)$ is a strictly convex function over $z > 0$. The second derivative of $f(z)$ is $\frac{1}{z^2}$. Thus

$$-\sum_{i=1}^n \gamma_i \log x_i \geq -\log \sum_{i=1}^n \gamma_i x_i$$

The proof follows by noting that log is an increasing function.

4. (10 points) Consider desiging the tube for holding *shuttle cocks.* It is essentially a cylinder of radius $r$ cm and height $h$ cm. The cost of painting the top and bottom of the cylinder is $cRs/cm^2$ and the cost of painting the sides is $dRs/cm^2$. How will you design the cyliner, that is choose $r$ and $h$, such that cost of painting it is minumum but it must hold 6 such shuttle cocks. In other words solve the following problem

$$\min_{r,h} 2\pi r^2 c + \pi r h d \quad \text{subject to } \pi r^2 h = V$$

where $V$ is the volume of space needed to hold the shuttle cocks. Use the previous question to solve the optimization problem.

**Solution:** Let $\gamma_1 z_1 = 2\pi r^2 c, \quad \gamma_2 z_2 = \pi r h d$ where $\gamma_1 + \gamma_2 = 1$ and $\gamma_1, \gamma_2 \geq 0$.

$$z_1^{\gamma_1} z_2^{\gamma_2} = A(\pi r^2 h)^a$$

By equating the exponents of $r$ and $h$ one obtains $2\gamma_1 + \gamma_2 = 2a$ and $\gamma_2 = a$. Thus $\gamma_1 = a/2$ and $\gamma_2 = a$. Since they should sum to 1, the value of $a = 2/3$. Using the previous question we deduce that

$$\gamma_1 z_1 + \gamma_2 z_2 \geq AV^{\frac{2}{3}}$$

and the minimum is attained for a choice of $r, h$ such that

$$z_1 = z_2 = AV^{\frac{2}{3}} \tag{1}$$

Plugging all these we obtain $z_1 = 6\pi r^2 c$ and $z_2 = \frac{3}{2}\pi r h d$ and the optimum cost

$$z_1^{\frac{a}{2}} z_2^{a} = 3(\frac{\pi}{2} c d^2)^{\frac{1}{3}} \left(\pi r^2 h\right)^{\frac{2}{3}}$$

$$z_1^{\frac{a}{2}} z_2^{a} = 3(\frac{\pi}{2} c d^2)^{\frac{1}{3}} V^{\frac{2}{3}}$$

Due to (1) the minimum is achieved at $r^3 = \frac{dV}{4\pi c}$ and $h^3 = \frac{16 c^2 V}{\pi d^2}$.

5. (10 points) **The case of sloppy stepsize:** Consider minimizing the function

$$f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} - \mathbf{b}^\top \mathbf{x}$$

over $\mathbf{x} \in \mathbb{R}^d$ with $Q \in \mathcal{S}_d^+, \mathbf{b} \in \mathbb{R}^d$ using the steepest descent iterates,

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - \alpha_k \nabla f(\mathbf{x}^{(k)})$$

executed with a *sloppy* step-size; each step-size $\alpha_k$ can be any element in $\{\alpha || \frac{\alpha}{\bar{\alpha}} - 1| \leq \delta\}$ The parameter $\bar{\alpha}$ is defined as the stepsize obtained through exact line search. How many iterations will it need to reach to find a point $\bar{\mathbf{x}}$ such that

$$f(\bar{\mathbf{x}}) - f^* \leq \epsilon$$

starting from an arbitrary point $\mathbf{x}^{(0)}$. Assume that you have chosen a stepsize which is farthest from $\bar{\alpha}$.

**Solution:** Consider the error function $E(\mathbf{x}) = f(\mathbf{x}) - f(\mathbf{x}^*) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^*)^\top Q(\mathbf{x} - \mathbf{x}^*)$. Recall that $E(\mathbf{x}) = \frac{1}{2}\nabla f(\mathbf{x})^\top Q^{-1}\nabla f(\mathbf{x})$. We will use $g_k$ to denote $\nabla f(\mathbf{x}^{(k)})$. From Taylor expansion one obtains that for any $\mathbf{x} = \mathbf{x}^{(k)} - \alpha g_k$

$$E(\mathbf{x}) = E(\mathbf{x}^{(k)}) - \alpha\|g_k\|^2 + \frac{1}{2}\alpha^2 g_k^\top Q g_k$$

Define $\bar{\alpha} = \frac{\|g_k\|^2}{g_k Q g_k}$ and completing squares lead to

$$E(\mathbf{x}) = E(\mathbf{x}^{(k)}) - \frac{1}{2}\bar{\alpha}^2 g_k^\top Q g_k + \frac{1}{2} g_k^\top Q g_k (\alpha - \bar{\alpha})^2$$

The decrease after each iteration is given by

$$E(\mathbf{x}^{(k)}) - E(\mathbf{x}^{(k+1)}) = \frac{1}{2} g_k^\top Q g_k \bar{\alpha}^2 - \frac{1}{2} g_k^\top Q g_k (\alpha - \bar{\alpha})^2 \geq \frac{1}{2} g_k^\top Q g_k \bar{\alpha}^2 (1 - \delta^2)$$

$$= \frac{1}{2}\frac{(\|g_k\|^2)^2}{g_k^\top Q g_k}(1 - \delta^2)$$

$$\frac{E(\mathbf{x}^{(k)}) - E(\mathbf{x}^{(k+1)})}{E(\mathbf{x}^{(k)})} \geq \frac{(\|g_k\|^2)^2}{(g_k^\top Q g_k)(g_k^\top Q^{-1} g_k}(1 - \delta^2)$$

Applying Kantorovch inequality we obatin

$$\frac{E(\mathbf{x}^{(k)}) - E(\mathbf{x}^{(k+1)})}{E(\mathbf{x}^{(k)})} \geq \frac{4\lambda_1\lambda_d}{(\lambda_1 + \lambda_d)^2}(1 - \delta^2)$$

Hence $\frac{E(\mathbf{x}^{(k+1)})}{E(\mathbf{x}^{(k)})} \leq r$ where $r = 1 - \frac{4\lambda_1\lambda_d}{(\lambda_1+\lambda_d)^2}(1 - \delta^2)$. From the above

$$E(\mathbf{x}^k) \leq r^k E(\mathbf{x}^0) \leq \epsilon$$

whenever $k \geq \frac{1}{\log\frac{1}{r}} \log \frac{E(\mathbf{x}^0)}{\epsilon}$

Check that $r = r_0 + (1 - r_0)\delta^2$ where $r_0 = \frac{(\lambda_1 - \lambda_d)^2}{\lambda_1 + \lambda_d)^2}$, is the rate obtained by using steepest descent.

6. (5 points) Consider $f$ as defined in Question 5. Show that Goldstein condition on the stepsize can be written as

$$\alpha \in \left\{ \alpha \middle| \left| \frac{\alpha}{\bar{\alpha}} - 1 \right| \leq \delta \right\}$$

Identify $\bar{\alpha}$ and $\delta$. Assume that the iterates are of the form

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \alpha_k \mathbf{u}$$

where $\mathbf{u}$ is a Descent direction at $\mathbf{x}^k$. Briefly comment the relative merit/demerit of Wolfe condition over this Goldstein in the context of this problem.

**Solution:** Introduce the following function

$$g(\alpha) = f(\mathbf{x}^{(k)} + \alpha \mathbf{u}) = g(0) + \alpha g'(0) + \frac{1}{2}\alpha^2 g"(0)$$

$$g'(0) = g_k^\top \mathbf{u}, \quad g"(0) = \mathbf{u}^\top Q \mathbf{u}$$

Define $\bar{\alpha} = -\frac{g'(0)}{g"(0)}$, the minimum of $g(\alpha)$, and

$$g(\alpha) = g(0) - \frac{1}{2}\bar{\alpha}^2 g"(0) + \frac{1}{2} g"(0)(\alpha - \bar{\alpha})^2$$

For any $0 \leq m \leq \frac{1}{2}$, the Goldstein condition can be stated as

$$g(0) - g"(0)(1-m)\bar{\alpha}\alpha \leq g(\alpha) \leq g(0) - g"(0) m \alpha \bar{\alpha}$$

where we have used the definition of $\bar{\alpha}$. This simplifies to

$$(1-m)\alpha\bar{\alpha} \geq \frac{1}{2}\bar{\alpha}^2 - \frac{1}{2}(\alpha - \bar{\alpha})^2 \geq m\alpha\bar{\alpha}$$

$$(1-m)\alpha\bar{\alpha} \geq -\frac{1}{2}\alpha^2 + \bar{\alpha}\alpha \geq m\alpha\bar{\alpha}$$

$$(1-m) \geq -\frac{1}{2}\frac{\alpha}{\bar{\alpha}} + 1 \geq m$$

$$2 - 2m \geq -\frac{\alpha}{\bar{\alpha}} + 2 \geq 2m$$

$$\delta \geq 1 - \frac{\alpha}{\bar{\alpha}} \geq -\delta$$

where we have used $1 - 2m = \delta$ and the proof follows.

As the optimum stepsize is already contained in the set there is little merit in using Wolfe condition over Goldstein.

**Rough Sheet 1**

**Rough Sheet 2**

**Rough Sheet 3**

**Rough Sheet 4**